

**REMARKS**

The Examiner rejected claims 17 under 35 U.S.C. §101 as being inoperative because of an inconsistency with the claims from which it depends. We have amended claim 17 to resolve this issue. More specifically, it now refers to the target document rather than plural documents.

The Examiner rejected claims 1-3, 10-14, 16 and 18 under 35 U.S.C. §112, 2<sup>nd</sup> paragraph, as being indefinite. We have amended the claims to address the Examiner's §112 concerns and to more particularly point out and distinctly claim the invention. We have also canceled claims 18 and 19.

After entering the amendments presented herein, claims 1-17, and 20 will be pending in this application.

We acknowledged the Examiner's indication that claims 14-16 would be allowable if rewritten to overcome the rejection(s) under 35 U.S.C. §112, 2<sup>nd</sup> paragraph, and to include all of the limitations of the base claim and any intervening claims.

The Examiner rejected claims 1, 3-12, 17 and 20 under 35 U.S.C. §102(b) as being anticipated by Smith et al. (Disambiguating Geographic Names in a Historical Digital Library) (a.k.a. Smith). We note, however, that Smith deals with one document at a time not a corpus of documents. The weights which he computes are not values that are derived from analysis of nor do they reflect a statistic for a large corpus of documents but rather each weight is derived from a single document and it is associated with that single document.

The language of the original claims, which is recited below, makes clear that the present claims are directed to deriving a statistics for a corpus of documents. The original claim recited:

...in a large corpus, identifying geo-textual correlations among readings of the toponyms within the plurality of toponyms

and

...using the identified geo-textual correlations to generate a value for a confidence that the selected toponym refers to a corresponding geographic location.

To remove any doubt that the referenced geo-textual correlations refer to a statistic that is derived from the corpus of documents, as opposed to a single document, we have amended claim 1 to recite:

...based on an analysis of all the documents within a large corpus of documents, identifying geo-textual correlations among readings of toponyms within the plurality of toponyms, wherein the geo-textual correlations are statistics derived for the corpus of documents rather than for any individual document within the corpus of documents

Smith simply does not have anything to do with identifying geo-textual correlations of the type recited in the claims.

The local nature of Smith's inquiry is clearly indicated in his description. He summarizes his methods as follows:

Our methods for performing these tasks [toponym disambiguation] rely on evidence that is internal or external to the text...Internal evidence includes the use of honorifics, generic geographical labels, or linguistic environment. External evidence includes gazetteers, biographical information, and generic linguistic knowledge. (page 5, line 37, to page 6, line 4).

Smith then describes how he scans the documents to find the toponyms and uses external sources (e.g. gazetteers) to find their geographical meaning. For the toponyms for which there is ambiguity, Smith uses a procedure to disambiguate them:

Disambiguating the possible place names then proceeds based on local context, document context, and general world knowledge. (emphasis added) (page 6, line 39).

And he describes each of these three procedures in greater detail in the following few paragraphs. His description makes clear that he is not generating a geo-textual correlation or any other statistic based on an analysis of a corpus of documents.

One of the advantages of the claimed approach (i.e., identifying geo-textual correlations) is that the resulting confidences have global significance; they reflect observations from a much larger

set of data; and they are more accurate measures of the relationship between toponym and place. Moreover, they can be further refined by taking to account the local context in which the particular toponym appears, as is described in the specification.

In support of his rejection of claim 10, the Examiner directs our attention to a passage on page 7 of Smith which he characterizes as disclosing boosting the value of a confidence. The paragraph reads as follows:

Each possible location for a toponym is given a score based on (a) its proximity to other toponyms around it, (b) its proximity to the centroid for the document, and (c) its relative importance – e.g. all other things being equal, nations get a higher score than cities. Also at this stage, the system discards as probably false positives places that lack an explicit disambiguator, that receive a low importance score, and that are far away from the local and document centroids. If not thus eliminated, the candidate toponym identification with the highest score is declared the winner. Once the work of the disambiguation system is done, the resulting toponyms are loaded into a relational database for access by the runtime digital library system.

But there is no mention in this paragraph of boosting a value of a confidence for a selected (toponym, place) pair. It simply describes how a score is computed. And though it does indicate that its score is based on the presence of other toponyms, it does not suggest that the selected (toponym, place) pair has a confidence value and that value is boosted by the presence of other toponyms.

We have amended claim 10 to make this distinction more clear. The amended claim 10 now recites in relevant part:

...obtaining a pre-computed initial value for the value of the confidence that the toponym of the selected (toponym, place) pair refers to the place of the selected (toponym, place) pair, said pre-computed initial value derived from a statistical observation about a large corpus of documents;

This makes clear that there is a pre-computed value for the confidence and it is that value that is boosted.

Application No. 10/803,831  
Amendment dated October 14, 2008  
Reply to Office Action of April 25, 2008

Docket No.: 0113744.00124US2

For at least the reasons stated above, we believe that the claims are in condition for allowance and therefore ask the Examiner to allow them to issue.

Please apply any charges not covered, or any credits, to Deposit Account No. 08-0219, under Order No. 0113744.00124US2 from which the undersigned is authorized to draw.

Respectfully submitted,

Dated: October 14, 2008



Eric L. Prahl  
Registration No.: 32,590  
Attorney for Applicant(s)

Wilmer Cutler Pickering Hale and Dorr LLP  
60 State Street  
Boston, Massachusetts 02109  
(617) 526-6000 (telephone)  
(617) 526-5000 (facsimile)